J. Appl. Numer. Optim. 8 (2026), No. 2, pp. 201-213
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.8.2026.2.03

A DUAL ALTERNATING DIRECTION METHOD OF MULTIPLIERS FOR TOEPLITZ-CONSTRAINED LOG-DETERMINANT OPTIMIZATION PROBLEMS

WENHAN JIA[1], CHENGJING WANG[1,*], PEIPEI TANG[2], AIMIN XU[3], JIAJIA WANG[1]

[1]*School of Mathematics, Southwest Jiaotong University, Chengdu 611731, China*
[2]*School of Computer and Computing Science, Hangzhou City University, Hangzhou 310015, China*
[3]*Institute of Mathematics, Zhejiang Wanli University, Ningbo 315100, China*

Abstract. In this paper, we study the log-determinant (log-det) optimization problem with a Toeplitz structure constraint, which has significant applications in sparse inverse covariance estimation. We propose a dual alternating direction method of multipliers (dADMM) to solve its dual formulation. Furthermore, we establish that the dADMM achieves global convergence under mild conditions. Numerical results demonstrate the efficacy and stability of the proposed algorithm.

Keywords. Alternating direction method of multipliers; Log-determinant optimization problem; Sparse inverse covariance estimation; Toeplitz matrix;

2020 Mathematics Subject Classification. 65K05, 90C06, 90C20.

1. INTRODUCTION

In this paper, by defining $\log 0 := -\infty$, we focus on the sparsity-regularized log-determinant (log-det) optimization problem with a Toeplitz structure constraint. Its primal-dual formulations are given by:

$$(P_0) \qquad \min_{X\in\mathbb{S}^n, x\in\mathbb{R}^{p+1}} \langle C,X\rangle - \mu\log\det(X) + \lambda\|x\|_1 \quad \text{s.t.} \quad \mathscr{T}x - X = 0,\ X \succeq 0,$$

$$(D_0) \qquad \min_{Y,Z\in\mathbb{S}^n, w\in\mathbb{R}^{p+1}} -\mu\log\det(Y) + \delta_{B_\infty^\lambda}(w) \quad \text{s.t.} \quad \mathscr{T}^*(Z) - w = 0,\ C - Y + Z = 0,\ Y \succeq 0,$$

where $C\in\mathbb{S}^n, \mu > 0, \lambda > 0$ are given parameters, the Toeplitz generator operator $\mathscr{T}:\mathbb{R}^{p+1}\to\mathbb{S}^n$ maps a vector $x\in\mathbb{R}^{p+1}$ to a linear combination of symmetric Toeplitz matrices:

$$\mathscr{T}x := \sum_{i=1}^{p+1} x_i T_i, \quad x\in\mathbb{R}^{p+1},$$

where $\{T_i\}_{i=1}^{p+1}$ is a basis set, each $T_i\in\mathbb{S}^n$ is a matrix with 1s on the i-th diagonal and 0s elsewhere. Its adjoint operator $\mathscr{T}^*:\mathbb{R}^{p+1}\to\mathbb{S}^n$ can be expressed as

$$\mathscr{T}^*X = \left[\langle T_1,X\rangle, \langle T_2,X\rangle, \cdots, \langle T_{p+1},X\rangle\right]^T.$$

*Corresponding author.
E-mail address: renascencewang@hotmail.com (C. Wang).
Received 22 July 2025; Accepted 4 November 2025; Published online 1 May 2026.

This class of optimization problems originates from parameter estimation in multivariate Gaussian distributions in statistics. The probability density function of a multivariate Gaussian distribution depends on the determinant of the covariance matrix Σ, and its log-likelihood function includes the term $\log\det(\Sigma)$. Given sample data $X_1,\ldots,X_n$, the maximum likelihood estimation (MLE) of the covariance matrix requires minimizing

$$\log\det(\Sigma)+\operatorname{tr}(\Sigma^{-1}S),$$

where S is the sample covariance matrix [1]. Early works by Fisher [2] and Hotelling [3] implicitly addressed the positive definiteness constraint and the parameter relationships for the covariance matrix through the log-determinant, though modern convex optimization terminology was not yet formalized. Subsequently, Shannon [4] introduced the mathematical theory of entropy, which established a rigorous connection between multivariate Gaussian entropy and the log-determinant term. Furthermore, the log-det term serves as a barrier function in optimization problems [5]. Today, log-det problems have broader applications in various fields.

Log-determinant (log-det) optimization serves as a fundamental framework for structured covariance matrix estimation in statistics, information theory, machine learning, and related disciplines. Advances in data acquisition technologies have elevated high-dimensional data analysis to a central task in applications such as financial time series forecasting, gene regulatory network inference, and wireless communication system identification. In these scenarios, the covariance matrix must not only capture complex dependencies between variables but also satisfy specific structural constraints to reflect inherent data patterns. Traditional sample covariance matrices face significant challenges in high dimensions: when the number of samples m is much smaller than the variable dimension n, the matrix singularity renders the inverse matrix nonexistent; even when $m>n$, sample covariance estimates exhibit high sensitivity to noise , leading to amplified estimation errors in off-diagonal elements [6]. This issue is particularly prominent in financial high-frequency trading data and single-cell genomics data, severely limiting model generalization.

To address these challenges, researchers have developed a framework combining regularization with structural constraints [7, 8]. Sparse regularization compresses redundant parameters, while structural constraints (Toeplitz, low-rank) leverage domain knowledge. The Toeplitz covariance structure naturally arises in many practical settings. It is especially valuable in signal processing [9] due to its capacity to capture time-shift invariance in stationary time series. For instance, in brain–computer interfaces, researchers have found that multichannel electroencephalogram covariance matrices can be well-modeled as block–Toeplitz under the assumption of short-window stationarity, which improves classification performance in linear discriminant analysis-based systems [10]. In multivariate time-series clustering, the Toeplitz inverse covariance-based clustering algorithm enforces a Toeplitz inverse covariance across short windows to capture time-invariant dependency networks [11].

A classical approach for the structured covariance matrix estimation is minimizing the Kullback-Leibler (KL) divergence. The KL divergence was first introduced by Kullback and Leibler [12] to measure differences between probability distributions. It is defined as

$$D_{\mathrm{KL}}(p\|q):=\frac{1}{2}\left[\operatorname{tr}(S^{-1}\Sigma)-\log\det(S^{-1}\Sigma)-n\right],$$

where n denotes the dimension, $p(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and $q(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, S)$ are the true and estimated distributions, respectively, with Σ being the true covariance matrix and S the estimated covariance matrix. This expression, sometimes called the entropy loss function or relative entropy, is one practical source of the log-det optimization problem's objective function. Later, Dhillon and Tropp [13] extended this framework to matrix spaces, applicable to general convex constraint sets. A special case is the log-det optimization problem with Toeplitz structure:

$$\min_{B \in \Omega} \left\{ \operatorname{tr}(A^{-1}B) - \log\det(A^{-1}B) \right\},$$

where $A \in \mathbb{S}^n$ is the empirical covariance matrix and Ω is the set of Toeplitz matrices:

$$\Omega = \left\{ \Theta \in \mathbb{S}^n_+ \,\middle|\, \Theta_{i,j} = 0 \text{ if } |i-j| > k \right\}.$$

Research has placed growing emphasis on increasingly focused on the estimation of covariance matrices Σ and their inverses Θ. Dempster [14] proposed the expectation-maximization algorithm, establishing a fundamental theoretical framework for covariance selection problems, particularly sparse covariance matrix estimation in Gaussian graphical models. Pan et al. [15] developed an innovative parameterization approach, expressing the covariance matrix as a polynomial function of autoregressive coefficients and innovation variances. Their method jointly optimized mean and covariance models using the Bayesian information criterion, offering an efficient solution for high-dimensional longitudinal data analysis. While this approach did not explicitly consider Toeplitz structures, its core contribution lay in computational complexity reduction through low-dimensional parameterization and alternating optimization strategies. Banerjee et al. [7] proposed the sparse maximum likelihood estimation method, which has significant influence in Gaussian graphical models and covariance selection. Building on this work, Yuan and Lin [8] constrained the sparsity of the precision matrix (inverse covariance) via the ℓ_1 regularization (or more complex non-convex penalties), leading to a convex optimization model known as the graphical Lasso:

$$\min_{\Theta \succeq 0} \left\{ \operatorname{tr}(S\Theta) - \log\det(\Theta) + \lambda \|\Theta\|_1 \right\}, \tag{1.1}$$

where $S \in \mathbb{S}^n$ is the empirical covariance matrix, and the regularization term $\lambda\|\Theta\|_1$ simultaneously enforces sparsity (revealing conditional independence) and preserves domain priors like the Toeplitz structure. This model maintains maximum likelihood estimation properties while achieving computational efficiency through sparsity, making it ideal for interpretable high-dimensional problems. In recent years, the covariance selection problems with more complicated regularization terms were proposed and numerical algorithms were designed to solve them, e.g., [16, 17, 18].

For a Toeplitz covariance matrix Σ, its inverse Θ typically exhibits banded sparsity, where non-zero elements concentrate near the diagonal. The bandwidth k reflects how quickly conditional dependencies decay with distance. Moreover, a zero of Θ_{ij} indicates variables i and j are conditionally independent. This banded structure implies local dependencies. For example, we consider an old-fashioned metronome: the current beat intensity is mainly influenced by the previous few beats, not by beats from the distant past. This "short-term memory" characteristic is precisely what the Toeplitz structure describes. We can also imagine a simple weather forecasting system studying the relationship between temperatures over three consecutive days.

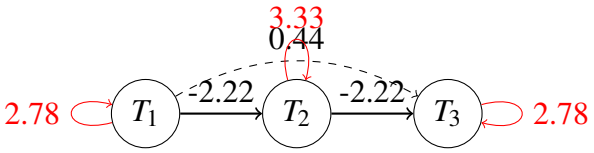

FIGURE 1. Conditional dependency network for temperature prediction (solid line: strong association, dashed line: weak association, red self-loop: self-regulation).

Suppose the temperatures satisfy the autoregressive relationship

$$T_t = 0.8T_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 0.36).$$

Its covariance matrix exhibits a Toeplitz structure

$$\Sigma = \begin{pmatrix} 1.0 & 0.8 & 0.64 \\ 0.8 & 1.0 & 0.8 \\ 0.64 & 0.8 & 1.0 \end{pmatrix}.$$

We observe the following features: the main diagonal elements are constant 1s (representing normalized temperature variance); the k-th sub-diagonal elements are ρ^ks ($\rho = 0.8$ is the decay coefficient); the symmetry ensures time-shift invariance. The corresponding precision matrix is

$$\Theta = \Sigma^{-1} = \begin{pmatrix} 2.78 & -2.22 & 0.44 \\ -2.22 & 3.33 & -2.22 \\ 0.44 & -2.22 & 2.78 \end{pmatrix}.$$

The main diagonal element 2.78 represents the temperature's self-regulation ability (positive influence of the current day's temperature on itself); the sub-diagonal negative element -2.22 indicates that an increase in yesterday's temperature suppresses today's temperature; and 0.44 can be seen as noise perturbation, suggesting no direct effect of the temperature two days prior on today's temperature. Figure 1 illustrates the conditional dependency network corresponding to this structure, showing direct connections only between adjacent nodes, aligning better with practical experience.

This simple model can describe other practical problems. For instance, the structure can represent a spring-mass system where each node is a spring scale, the main diagonal positive terms represent self-stiffness, the off-diagonal negative terms represent interactions between adjacent springs, and zero elements indicate no connection between non-adjacent nodes. In MIMO systems, the received signal can be modeled as $y = Hx + n$, where the channel matrix H exhibits a block-Toeplitz structure composed of submatrices H_{ij}, each representing a Toeplitz pulse shaping matrix G_l from multipath components (see [19, Eq. (2)-(3)]).

The Toeplitz constraint offers two advantages for our problems. Practically, it aligns with the principle of local interaction prevalent in many systems. For example, current signals in wireless communication are mainly affected by recent interference, tomorrow's weather likely depends primarily on today's state. Computationally, the number of independent parameters decreases from $O(n^2)$ for a general matrix to $O(n)$ for a Toeplitz matrix, significantly lowering algorithm design difficulty and improving computational efficiency.

However, existing algorithms face significant computational challenges when simultaneously addressing sparsity and Toeplitz constraints. Traditional Newton methods require repeated projections to ensure positive definiteness, with the Hessian matrix computation complexity scaling as $O(n^6)$; first-order methods are scalable but struggle to balance constraint violation and convergence speed [20]. While Boyd et al. [1] systematically proposed the ADMM framework for sparsity-regularized problems, enabling distributed computing for large-scale sparse optimization, its convergence speed depends critically on empirically tuned parameters. Wang et al. [21] proposed a Newton-conjugate gradient primal proximal point algorithm for log-det problems, especially efficient in the maximum likelihood sparse estimation for Gaussian graphical models. Lin et al. [20] introduced entropy loss functions for Toeplitz constraints but were limited by inexact Newton steps and high-dimensional Hessian computation bottlenecks, only solving small-scale problems. To avoid the computational bottlenecks associated with directly enforcing Toeplitz structural constraints and projection operations, we propose a dual alternating direction method of multipliers (dADMM) framework. Furthermore, we establish the global convergence of dADMM under mild conditions, providing a rigorous theoretical foundation. Comprehensive numerical experiments demonstrate the efficacy of our proposed dADMM approach, including comparative evaluations with the primal alternating direction method of multipliers (pADMM).

The remainder of this paper is organized as follows. Section 2 introduces the necessary mathematical preliminaries and theoretical foundations for the dADMM. Section 3 presents the complete algorithmic framework of our proposed dADMM approach, including implementation details and computational considerations. In Section 4, we establish the convergence analysis of the dADMM algorithm. Section 5 demonstrates the numerical performance of our method through comprehensive experiments and comparative analysis. Finally, Section 6 concludes the paper with a summary of key findings and potential directions for future research.

1.1. **Notation.** In this paper, we assume that all vectors are finite-dimensional. $\mathbb{R}^n$ denotes n-dimensional Euclidean space, $\mathbb{R}^{m\times n}$ is the set of all $m\times n$ real matrices, $\mathbb{S}^n$ is the space of symmetric matrices. $\langle\cdot,\cdot\rangle$ denotes the standard trace inner product on $\mathbb{R}^{m\times n}$. For matrices, the notation $\|\cdot\|$ without subscript denotes the Frobenius norm, defined as $\|A\|_F := \sqrt{\langle A,A\rangle}$ for any $A\in\mathbb{R}^{m\times n}$. For vectors, $\|\cdot\|$ without subscript denotes the Euclidean norm $\|\cdot\|_2$. Other norms (e.g., $\|\cdot\|_1$, $\|\cdot\|_\infty$) will always carry explicit subscripts.

The ℓ_∞-norm ball of radius $\lambda>0$ is denoted as $B_\infty^\lambda = \{w \mid \|w\|_\infty \le \lambda\}$. For any convex set C, the indicator function $\delta_C(w)$ is

$$\delta_C(w) = \begin{cases} 0, & \text{if } w\in C,\\ \infty, & \text{otherwise.}\end{cases}$$

Let $C\subseteq\mathbb{R}^n$ be a non-empty closed convex set. The projection operator $\Pi_C:\mathbb{R}^n\to C$ is defined as

$$\Pi_C(x) := \arg\min_{y\in C}\|y-x\|.$$

This operator is well-defined and single-valued due to the properties of closed convex sets.

2. PRELIMINARIES

To establish the framework for our algorithm, we provide some necessary background knowledge in this section.

In optimization problems, especially those with non-smooth regularization terms (e.g., the ℓ_1-norm), the proximal operator plays a crucial role. It allows us to handle non-smooth terms efficiently by transforming them into a minimization problem that often has a closed-form solution.

Definition 2.1. [22] Given a proper lower semicontinuous convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and parameter $\lambda > 0$, its **proximal operator** is defined as

$$\mathrm{Prox}_{\lambda f}(v) := \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda}\|x - v\|^2 \right\}.$$

In the subproblems of the log-det optimization , we need to solve a minimization problem that combines a least squares term with the log-det one. The following lemma provides an explicit solution to these subproblems, which is critical for updating the variable Y in (D_0) of our dADMM.

Lemma 2.1. [21] *Let Y be an $n \times n$ symmetric matrix with the eigen-decomposition $Y = PDP^T$, $D = diag(d)$, and assume $d_1 \geq d_2 \geq \cdots \geq d_r > 0 \geq d_{r+1} \geq \cdots \geq d_n$. Let $\gamma > 0$, and define two scalar functions*

$$\Phi_\gamma^+(x) := \frac{\sqrt{x^2 + 4\gamma} + x}{2}, \quad \Phi_\gamma^-(x) := \frac{\sqrt{x^2 + 4\gamma} - x}{2}, \quad \forall x \in \mathbb{R}.$$

Their matrix counterparts are

$$\Phi_\gamma^+(Y) := P\Phi_\gamma^+(diag(d))P^T, \quad \Phi_\gamma^-(Y) := P\Phi_\gamma^-(diag(d))P^T.$$

Then

$$Y = \Phi_\gamma^+(Y) - \Phi_\gamma^-(Y), \quad \textit{with} \quad \Phi_\gamma^+(Y), \Phi_\gamma^-(Y) \succ \mathbf{0} \quad \textit{and} \quad \Phi_\gamma^+(Y)\Phi_\gamma^-(Y) = \gamma I$$

and the solution

$$Z^* = \arg\min_{Z \succ \mathbf{0}} \left\{ \frac{1}{2\lambda}\|Y - Z\|^2 - \mu \log\det(Z) \right\} = \Phi_\gamma^+(Y),$$

and

$$\frac{1}{2\lambda}\|Y - Z^*\|^2 - \mu \log\det(Z^*) = \frac{1}{2\lambda}\|\Phi_\gamma^-(Y)\|^2 - \mu \log\det(\Phi_\gamma^+(Y)),$$

where $\gamma = \lambda\mu$.

3. ALGORITHM

In this section, we derive the dADMM for the Toeplitz-constrained log-det problem.

The Karush-Kuhn-Tucker (KKT) condition for problem (P_0) is

$$\begin{cases} C-Y+Z=0, \\ \mathscr{T}^*(Z)-w=0, \\ XY-\mu I=0, \\ \mathscr{T}x-X=0, \\ w-\mathrm{Prox}_{\sigma\lambda\|\cdot\|_1}(x+\frac{w}{\sigma})=0, \\ X\succeq 0, Y\succeq 0. \end{cases}$$

The augmented Lagrangian function for problem (D_0) with a parameter $\sigma>0$ is

$$\begin{aligned}\mathscr{L}_\sigma(Y,Z,w;x,X) = &-\mu\log\det(Y)+\delta_{B_\infty^\lambda}(w)-\langle x,\mathscr{T}^*(Z)-w\rangle-\langle X,C-Y+Z\rangle \\ &+\frac{\sigma}{2}\|\mathscr{T}^*(Z)-w\|^2+\frac{\sigma}{2}\|C-Y+Z\|^2.\end{aligned}$$

At the k-th iteration of the dADMM, we need to minimize $\mathscr{L}_\sigma(Y^k,Z^k,w^k;x^k,X^k)$ with respect to the variables (Y,w) and Z alternatingly. The key details are as follows:

• For the variables (Y^{k+1},w^{k+1}):

$$\begin{aligned}Y^{k+1} &= \arg\min_{Y\succeq 0}\left\{-\mu\log\det(Y)+\frac{\sigma}{2}\|Y-(C+Z^k-\tfrac{1}{\sigma}X^k)\|^2\right\} \\ &= \Phi^+_{\frac{\mu}{\sigma}}\left(C+Z^k-\tfrac{1}{\sigma}X^k\right) \quad \text{(by Lemma 2.2)},\end{aligned}$$

and

$$\begin{aligned}w^{k+1} &= \arg\min_{w}\left\{\langle x^k,w\rangle+\frac{\sigma}{2}\|\mathscr{T}^*(Z^k)-w\|^2+\delta_{B_\infty^\lambda}(w)\right\} \\ &= \arg\min_{w\in B_\infty^\lambda}\left\{\|w-(\mathscr{T}^*(Z^k)-\frac{1}{\sigma}x^k)\|^2\right\} \\ &= \Pi_{B_\infty^\lambda}\left(\mathscr{T}^*(Z^k)-\tfrac{1}{\sigma}x^k\right).\end{aligned}$$

• For the variable Z^{k+1}:

$$\begin{aligned}Z^{k+1} &= \arg\min_{Z}\left\{-\langle x^k,\mathscr{T}^*(Z)\rangle-\langle X^k,Z\rangle+\frac{\sigma}{2}\|\mathscr{T}^*(Z)-w^{k+1}\|^2+\frac{\sigma}{2}\|C-Y^{k+1}+Z\|^2\right\} \\ &= (\mathscr{T}\mathscr{T}^*+\mathscr{I})^{-1}\left(Y^{k+1}-C+\frac{1}{\sigma}(\mathscr{T}x^k+X^k)+\mathscr{T}w^{k+1}\right).\end{aligned}$$

In practical computation, we apply the preconditioned conjugate gradient method to obtain Z^{k+1}.

In the following, we provide the algorithm framework of the dADMM for problem (D_0).

Algorithm 1 (dADMM for (D_0)):
Given initial values Y^0, Z^0, w^0, x^0, X^0, the penalty parameter $\sigma > 0$, step size $\tau \in (0, \frac{1+\sqrt{5}}{2})$, and stopping criterion. Let $k = 0, 1, \cdots$, and perform the following iterations:
Step 1: Compute (Y^{k+1}, w^{k+1}):

$$Y^{k+1} = \Phi^{+}_{\frac{\mu}{\sigma}}\left(C + Z^k - \frac{1}{\sigma}X^k\right),\ w^{k+1} = \Pi_{B^{\lambda}_{\infty}}\left(\mathscr{T}^*(Z^k) - \frac{1}{\sigma}x^k\right).$$

Step 2: Compute Z^{k+1} by $Z^{k+1} = (\mathscr{T}\mathscr{T}^* + \mathscr{I})^{-1}\left(Y^{k+1} - C + \frac{1}{\sigma}(\mathscr{T}x^k + X^k) + \mathscr{T}w^{k+1}\right)$.
Step 3: Update the multipliers x^{k+1}, X^{k+1}:

$$x^{k+1} = x^k - \tau\sigma\left(\mathscr{T}^*(Z^{k+1}) - w^{k+1}\right),\ X^{k+1} = X^k - \tau\sigma\left(C - Y^{k+1} + Z^{k+1}\right).$$

Step 4: If the stopping criterion is satisfied, terminate; otherwise, set $k \leftarrow k+1$ and go to Step 1.

4. Convergence Analysis

In this section, we present the convergence results for the dADMM applied to problem (D_0). For technical reasons, consider the constraint qualification :

CQ: There exists $(X^0, x^0, Y^0, Z^0, w^0) \in \mathbb{S}^n_{++} \times \mathbb{R}^{p+1} \times \mathbb{S}^n_{++} \times \mathbb{S}^n \times \mathbb{R}^{p+1} \cap \mathbb{P}$, where $\mathbb{P}$ is the constraint set of (P_0) and (D_0).

Denote $\mathscr{M} := \mathbb{S}^n \times \mathbb{R}^{p+1} \times \mathbb{S}^n \times \mathbb{S}^n \times \mathbb{R}^{p+1}$. Then the KKT residual mapping $\mathscr{R} : \mathscr{M} \to \mathscr{M}$ is given by

$$\mathscr{R}(m) := \begin{bmatrix} C - Y + Z \\ \mathscr{T}^*(Z) - w \\ XY - \mu I \\ \mathscr{T}x - X \\ w - \mathrm{Prox}_{\sigma\lambda\|\cdot\|_1}(x + w/\sigma) \end{bmatrix}, \quad \forall m = (X, x, Y, Z, w) \in \mathscr{M}.$$

A point $m^* \in \mathscr{M}$ is a KKT point iff $\mathscr{R}(m^*) = 0$ with $X \succeq 0$, $Y \succeq 0$.

Based on [23, Theorem B.1], we have the following result.

Theorem 4.1. *(Global Convergence) Let the solution sets of (P_0) and (D_0) be non-empty and step size $\tau \in (0, (1+\sqrt{5})/2)$. Then the sequence $\{(Y^k, Z^k, w^k)\}$ generated by the dADMM converges to the optimal solution of problem (D_0), and $\{(X^k, x^k)\}$ converges to the optimal solution of problem (P_0).*

Proof. Under the **CQ**, by [22, Corollary 28.2.2 and 28.3.1], $(\overline{Y}, \overline{Z}, \overline{w})$ is an optimal solution to problem (D_0) iff there exist multipliers $(\overline{X}, \overline{x})$ such that

$$\begin{cases} C - \overline{Y} + \overline{Z} = 0, \\ \mathscr{T}^*(\overline{Z}) - \overline{w} = 0, \\ \overline{X}\overline{Y} - \mu I = 0, \\ \mathscr{T}\overline{x} - \overline{X} = 0, \\ \overline{w} - \mathrm{Prox}_{\sigma\lambda\|\cdot\|_1}(\overline{x} + \overline{w}/\sigma) = 0, \\ \overline{X} \succeq 0, \overline{Y} \succeq 0. \end{cases}$$

Since $\begin{bmatrix} \sigma\mathscr{I} & 0 \\ 0 & \sigma\mathscr{I} \end{bmatrix} \succ 0$ and $\mathscr{T}\mathscr{T}^* + \mathscr{I} \succ 0$, we obtain by [23, Theorem B.1] the desired result immediately. □

5. Numerical Experiments

In this section, we validate dADMM's effectiveness through numerical experiments on both synthetic and real datasets. All experiments are performed in MATLAB 2023a on a Windows 10 desktop (Intel i5-13600KF, 3.50 GHz, 16GB RAM).

5.1. **Stopping Criteria.** In our experiments, we evaluate the solution quality based on the following measurements

$$\begin{aligned}
\text{pobj} &= \langle C, X\rangle - \mu \log\det(X) + \lambda\|x\|_1, \\
\text{dobj} &= -\mu \log\det(Y) + \delta_{B_\infty^\lambda}(w), \\
R_p &= \frac{\|\mathscr{T}x - X\| + \|y - x\|}{\max\{1, \|\mathscr{T}x\|, \|X\|, \|x\|, \|y\|\}}, \\
R_d &= \frac{\|C + Y - Z\| + \|\mathscr{T}^*(Z) - w\|}{\max\{1, \|C\|, \|Z\|, \|w\|\}}, \\
R_c &= \frac{\|XY - \mu I\|}{\max\{1, \|X\|, \|Y\|\}}, \\
R_g &= \frac{|\text{pobj} - \text{dobj}|}{1 + |\text{pobj}| + |\text{dobj}|},
\end{aligned}$$

where pobj and dobj are primal and dual objectives; R_p, R_d, R_c, R_g represent the primal infeasibility, dual infeasibility, complementary slackness violation, and duality gap. The iteration process stops when $\eta_{KKT} := \max\{R_p, R_d, R_c\} \le \texttt{tol} = 10^{-6}$ or the maximum iterations $\texttt{maxiter} = 10000$ is reached.

5.2. **pADMM.** For the comparison with the dADMM, we describe the details of the pADMM.

Algorithm 2 (pADMM for (D_0)):
Given initial values X^0, x^0, y^0, Z^0, w^0, penalty parameter $\sigma > 0$, step size $\tau \in (0, \frac{1+\sqrt{5}}{2})$, and stopping criterion. Let $k = 0, 1, \cdots$, and perform the following iterations:
Step 1: Compute (X^{k+1}, y^{k+1}):

$$X^{k+1} = \Phi^+_{\frac{\mu}{\sigma}}\left(\mathscr{T}x^k - \frac{1}{\sigma}(C + Z^k)\right),\ y^{k+1} = \text{Prox}_{\frac{\lambda}{\sigma}\|\cdot\|_1}\left(x^k + \frac{1}{\sigma}w^k\right).$$

Step 2: Compute x^{k+1} by $x^{k+1} = (\mathscr{T}^*\mathscr{T} + I)^{-1}\left(\mathscr{T}^*(X^{k+1}) + y^{k+1} + \frac{\mathscr{T}^*(Z^k) - w^k}{\sigma}\right)$.
Step 3: Update the multipliers Z^{k+1}, w^{k+1}:

$$Z^{k+1} = Z^k - \tau\sigma(\mathscr{T}x^{k+1} - X^{k+1}), w^{k+1} = w^k - \tau\sigma(y^{k+1} - x^{k+1}).$$

Step 4: If the stopping criterion is satisfied, terminate; otherwise, set $k \leftarrow k + 1$ and go to Step 1.

5.3. **Numerical Results on Random Data.** In this part, we compare the performances of the dADMM and pADMM on randomly generated datasets. The sample covariance matrix C in (P_0) is generated randomly. First, we draw k independent samples from $\mathcal{N}(0,\Sigma)$ to construct a data matrix $D \in \mathbb{R}^{k\times n}$. Next, we compute the empirical covariance $\widehat{C} = \mathrm{cov}(D)$. For the scale normalization, we then derive the scaling matrix $\Delta = \mathrm{diag}(1/\sqrt{\mathrm{diag}(\widehat{C})})$ and finally obtain the target matrix via $C = \Delta\widehat{C}\Delta$. The regularization parameter λ is selected via the 5-fold cross-validation from a set of logarithmically spaced candidates.

TABLE 1. pADMM vs. dADMM on Random Data

(n,p)	λ	algorithm	iter	η_{KKT}	R_g	obj	time (s)
(200,40)	1.00e-04	pADMM	42	8.30e-07	3.40e-08	4.4125e+02	0.10
		dADMM	34	7.10e-07	8.90e-08	4.4125e+02	0.10
(350,70)	1.45e-03	pADMM	33	8.11e-07	8.50e-07	5.5309e+02	0.24
		dADMM	46	9.19e-07	6.50e-07	5.5309e+02	0.43
(500,100)	8.50e-02	pADMM	60	9.22e-07	4.00e-10	5.7718e+02	0.55
		dADMM	37	6.05e-07	2.00e-07	5.7718e+02	0.95
(750,150)	2.66e-02	pADMM	31	8.09e-07	5.80e-08	1.1349e+03	1.47
		dADMM	32	9.63e-07	1.60e-08	1.1349e+03	1.96
(1000,200)	1.32e-02	pADMM	114	9.11e-07	1.30e-05	1.8351e+03	11.45
		dADMM	57	7.99e-07	1.90e-06	1.8351e+03	8.84
(1200,240)	1.39e-02	pADMM	164	9.98e-07	1.20e-04	-3.7478e+01	23.81
		dADMM	71	5.92e-07	5.80e-06	-3.7478e+01	7.32
(1400,280)	8.69e-03	pADMM	216	9.79e-07	2.70e-05	-2.6322e+02	46.99
		dADMM	42	5.92e-07	8.80e-07	-2.6322e+02	13.59
(1600,320)	6.87e-03	pADMM	286	9.69e-07	2.00e-05	-5.5650e+02	93.53
		dADMM	30	7.84e-07	7.90e-07	-5.5650e+02	15.43
(1800,360)	1.10e-02	pADMM	361	9.82e-07	2.00e-05	-8.6660e+02	137.45
		dADMM	31	9.53e-07	9.60e-08	-8.6660e+02	17.12
(2000,400)	1.76e-02	pADMM	446	9.96e-07	2.00e-05	-1.2059e+03	225.64
		dADMM	39	7.38e-07	3.20e-07	-1.2059e+03	27.97

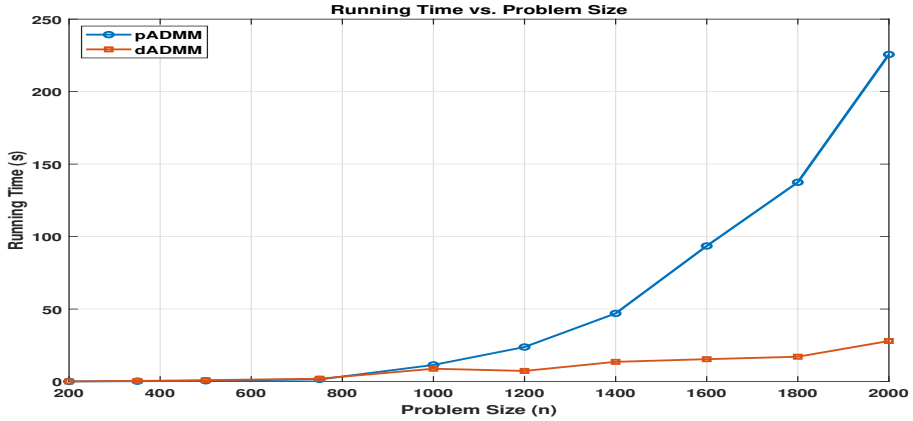

FIGURE 2. Running time vs. problem size.

In Table 1, we document the computational results of the pADMM and dADMM on synthetically generated datasets. In the table, we report the problem size ((n,p)), the regularization parameter (λ), the iteration number (iter), the KKT residual (η_{KKT}), the duality gap (R_g), the primal objective function value (obj) and the computing time (time) in seconds. For simplicity, the exponential notation adopts the convention "a.bec" denoting $a.b \times 10^c$ (e.g., 8.30e-07 signifies 8.30×10^{-7}).

We can see from the table, for small-scale problems (e.g., $n \leq 750$), the pADMM and dADMM exhibit comparable computational efficiency. However, when the scale becomes large, the dADMM obviously outperforms the pADMM. Figure 2 further corroborates dADMM's superior scalability, showing markedly better performance than the pADMM as the scale increases, underscoring its robustness for high-dimensional problems.

5.4. **Numerical Results on UCI Datasets.** In this part, we conduct experiments comparing the performances of the dADMM and pADMM across diverse UCI datasets featuring the structured covariance constraints. The dataset characteristics and numerical results are detailed below.

LSVT Voice Rehabilitation: This medical rehabilitation dataset ($n = 309$ voice features, we set $p = 56$ time windows) contains time-series voice measurements exhibiting potential stationarity properties. Capitalizing on the hypothesized time-shift invariance in vocal dynamics, we incorporate a Toeplitz structural constraint to model temporal correlations in vocal muscle activation patterns. The optimization framework identifies a regularized precision matrix X capturing feature interdependencies under this structural assumption, potentially revealing clinically relevant biofeature combinations.

Madelon: As a NIPS 2003 feature selection benchmark ($n = 500$ artificial features), this high-dimensional classification challenge contains irrelevant and redundant variables. To assess our model's capacity for capturing structured dependencies, we introduce an artificial Toeplitz covariance structure modeling distance-decaying feature correlations. This synthetic framework specifically evaluates the algorithm's ability to recover exponentially decaying dependencies while filtering noise in feature-rich environments.

ISOLET: In this speech recognition dataset ($n = 617$ Mel-frequency cepstral coefficients, $7,800$ samples), short-term acoustic features may exhibit correlation structures. Reflecting the spectral continuity between neighboring coefficients, we apply a Toeplitz constraint to model hypothesized decay patterns. This regularization aims to capture physical continuities in vocal tract resonances.

MicroMass: Comprising mass spectrometry measurements ($n = 1300$ m/z channels, 931 samples), this microbial identification dataset exhibits isotopic periodicity. Drawing on the physical properties of ^{13}C isotope envelopes, we implement a Toeplitz kernel to represent expected correlation decay between adjacent mass-to-charge channels.

GISETTE: This handwritten digit recognition dataset ($n = 1500$ feature-selected pixels) contains spatial image data with geometric correlations. Informed by prior knowledge of stroke continuity, we employ a Toeplitz constraint to approximate spatial decay in pixel covariance, modeling physical writing stroke continuities.

The results in Table 2 confirm that both algorithms attain the solutions with nearly the same quality. In general, the dADMM is more efficient.

TABLE 2. pADMM vs. dADMM on UCI Datasets

(n,p)	λ	algorithm	iter	η_{KKT}	R_g	obj	time (s)
(309,56)	1.62e+01	pADMM	71	8.40e-07	2.30e-07	5.1258e+02	0.34
		dADMM	55	7.45e-07	4.10e-07	5.1258e+02	0.34
(500,100)	1.00e-04	pADMM	29	6.51e-07	1.30e-07	4.9992e+02	0.52
		dADMM	30	7.73e-07	4.00e-07	4.9992e+02	0.74
(617,114)	1.95e+01	pADMM	75	8.79e-07	4.00e-07	9.7522e+02	2.25
		dADMM	65	7.51e-07	4.00e-07	9.7522e+02	2.63
(1300,259)	1.00e-04	pADMM	62	9.90e-07	1.20e-05	1.1898e+03	14.75
		dADMM	33	9.97e-07	5.20e-07	1.1898e+03	5.82
(1500,100)	1.14e-02	pADMM	44	6.51e-07	3.00e-07	2.9821e+03	15.26
		dADMM	36	9.12e-07	1.40e-09	2.9821e+03	8.51

6. CONCLUSION

In this paper, we adopted a dADMM for the Toeplitz-constrained log-det optimization problem, proved its global convergence under certain assumptions. The numerical experiments validated its effectiveness and stability. To improve the efficiency of computing, we plan to design a more efficient algorithm to solve this problem.

Acknowledgements

We sincerely thank the Editor-in-Chief Professor Xiaolong Qin and the anonymous reviewer for their valuable comments, which greatly improved the quality of this paper. Chengjing Wang's work was supported in part by Zhejiang Provincial Natural Science Foundation of China (Grant No. LMS26A010022), the National Natural Science Foundation of China (No. U21A20169). Peipei Tang's work was partly supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LMS26A010022). Aimin Xu's work was supported in part by Zhejiang Provincial Natural Science Foundation of China (Grant No. LTGY23H240002).

REFERENCES

[1] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.
[2] R. A. Fisher, Theory of statistical estimation, Math. Proc. Cambridge Philos. Soc. 22 (1925), 700–725.
[3] H. Hotelling, Relations between two sets of variates, Biometrika. 28 (1936), 321–377.
[4] C. E. Shannon, A mathematical theory of communication, Bell Sys. Tech. J. 27 (1948), 379–423.
[5] Y. Nesterov, A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, SIAM, Philadelphia, 1994.
[6] P. J. Bickel, E. Levina, Covariance regularization by thresholding, Ann. Statist. 36 (2008), 2577–2604.
[7] O. Banerjee, L. El Ghaoui, A. d'Aspremont, Model selection through sparse maximum likelihood estimation, J. Mach. Learn. Resh. 9 (2008), 485–516.
[8] M. Yuan, Y. Lin, Model selection and estimation in the Gaussian graphical model, Biometrika. 94 (2007), 19–35.
[9] D. Cederberg, Toeplitz covariance estimation with applications to music, Signal Process. 221 (2024), 109506.
[10] J. Sosulski, M. Tangermann, Introducing block-toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain-computer interfaces, J. Neural Eng. 19 (2022), 066014.

[11] D. Hallac, S. Vare, S. Boyd, J. Leskovec, Toeplitz inverse covariance-based clustering of multivariate time series data, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 215–223, 2017.
[12] S. Kullback, R. A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951), 79–86.
[13] I. S. Dhillon, J. A. Tropp, Matrix nearness problems with Bregman divergences, SIAM J. Matrix Anal. Appl. 29 (2007), 1120–1146.
[14] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. R. Stat. Soc. Ser. B. Stat. Methodol. 39 (1977), 1–38.
[15] J. Pan, G. Mackenzie, On modelling mean-covariance structures in longitudinal studies, Biometrika. 90 (2003), 239–244.
[16] T. Nakagaki, M. Fukuda, S. Kim, M. Yamashita, A dual spectral projected gradient method for log-determinant semidefinite problems, Comput. Optim. Appl. 76 (2020), 33–68.
[17] M. Lin, D. Sun, K. C. Toh, C. Wang, Estimation of sparse Gaussian graphical models with hidden clustering structure, J. Mach. Learn. Res. 25 (2024), 1–36.
[18] C. Wang, P. Tang, A dual semismooth newton based augmented Lagrangian method for large-scale linearly constrained sparse group square-root Lasso problems, J. Sci. Comput. 96 (2023), 45.
[19] G. G. Raleigh, J. M. Cioffi, Spatio-temporal coding for wireless communication, IEEE Trans. Commun. 46 (1998), 357–366.
[20] L. Lin, N. J. Higham, J. Pan, Covariance structure regularization via entropy loss function, Comput. Stat. Data Anal. 72 (2014), 315–327.
[21] C. Wang, D. Sun, K. C. Toh, Solving log-determinant optimization problems by a Newton-CG primal proximal point algorithm, SIAM J. Optim. 20 (2010), 2994–3013.
[22] R. T. Rockafellar, Convex Analysis, Princeton University Press, New Jersey, 1970.
[23] M. Fazel, T. K. Pong, D. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl. 34 (2013), 946–977.